Accepted for publication on 03/06/2020 in
Proceedings of the National Academy of Sciences USA

PNAS MS# 2019-18797 - 2nd Revision

Ancestral regulatory mechanisms specify conserved midbrain circuitry in arthropods and vertebrates

Jessika C. Bridi[1,†], Zoe N. Ludlow[1,†], Benjamin Kottler[1], Beate Hartmann[2], Lies Vanden Broeck[3], Jonah Dearlove[1], Markus Göker[4], Nicholas J. Strausfeld[5], Patrick Callaerts[3] and Frank Hirth[1,*]

[1]King's College London, Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom;
[2]University of Basel; Switzerland;
[3]Laboratory of Behavioural & Developmental Genetics, Department of Human Genetics, KU Leuven - University of Leuven, Leuven, Belgium;
[4]Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany;
[5]Department of Neuroscience and Center for Insect Sciences, University of Arizona, Tucson, USA.

[†]These authors contributed equally to this work.

Correspondence to:
Dr. Frank Hirth, King's College London, Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, Cutcombe Road, SE5 9RX, London, United Kingdom; Tel: ++44 20 7848 0786; email: Frank.Hirth@kcl.ac.uk

ABSTRACT

Corresponding attributes of neural development and function suggest arthropod and vertebrate brains may have an evolutionarily conserved organization. However, the underlying mechanisms have remained elusive. Here we identify a gene regulatory and character identity network defining the deutocerebral-tritocerebral boundary (DTB) in *Drosophila*. This network comprises genes homologous to those directing midbrain-hindbrain boundary (MHB) formation in vertebrates and their closest chordate relatives. Genetic tracing reveals that the embryonic DTB gives rise to adult midbrain circuits that in flies control auditory and vestibular information processing and motor coordination, as do MHB-derived circuits in vertebrates. DTB-specific gene expression and function is directed by cis-regulatory elements of developmental control genes that include homologs of mammalian *Zinc finger of the cerebellum* and *Purkinje cell protein 4*. *Drosophila* DTB-specific cis-regulatory elements correspond to regulatory sequences of human *ENGRAILED-2, PAX-2* and *DACHSHUND-1* that direct MHB-specific expression in the embryonic mouse brain. We show that cis-regulatory elements and the gene networks they regulate, direct the formation and function of midbrain circuits for balance and motor coordination in insects and mammals. Regulatory mechanisms mediating the genetic specification of cephalic neural circuits in arthropods correspond to those in chordates, thereby implying their origin before the divergence of deuterostomes and ecdysozoans.

Keywords: Brain; evolution; neural circuit; gene regulatory network; homology.

SIGNIFICANCE

Comparative developmental genetics indicate insect and mammalian forebrains form and function in comparable ways. However, these data are open to opposing interpretations that advocate either a single origin of the brain and its adaptive modification during animal evolution; or multiple, independent origins of the many different brains present in extant Bilateria. Here, we describe conserved regulatory elements that mediate the spatiotemporal expression of developmental control genes directing the formation and function of midbrain circuits in flies, mice and human. These circuits develop from corresponding midbrain-hindbrain boundary regions and regulate comparable behaviors for balance and motor control. Our findings suggest that conserved regulatory mechanisms specify cephalic circuits for sensory integration and coordinated behaviour common to all animals that possess a brain.

\body

Many components of the vertebrate and arthropod forebrain share features regarding neural arrangements along the rostrocaudal axis and their connections of sensory and motor pathways with higher integrative centres. This is exemplified in ancestral vertebrate lineages from which lampreys and hagfish derived. Rostral neuropils of the forebrain encode visual and olfactory information that are relayed to further forebrain centers integrating these modalities (1,2). Circuits involved in vestibular reception and integration, and by extension acoustic perception, develop in more caudal territories of the anterior brain that arise in the telencephalon of vertebrates (3). Corresponding arrangements are found in the deutocerebrum of arthropods (4), amongst which Onychophora offer comparable proxies of ancestral neural arrangements (5).

The formation of these neural arrangements is mediated by conserved developmental control genes acting along anterior-posterior (AP) and dorso-ventral (DV) axes of the embryonic nervous system (6-13). For example, the *Drosophila* gene *orthodenticle (otd)* and its mammalian *Otx* homologues are required in both for rostral brain development (14,15). In cross-phylum rescue experiments, human *OTX2* restores fly brain formation in *otd* mutant embryos (16,17) while fly *otd* can replace *Otx1/2* in mouse head and forebrain formation (18,19). Fly *engrailed* can replace *Engrailed-1* in mouse midbrain-hindbrain boundary (MHB) development (20). Further cross-phyletic studies revealed correspondences in developmental genetic mechanisms underlying circuit formation and information processing of the vertebrate basal ganglia and the arthropod central complex, including pathologies (9,21-23). These similarities also extend to comparisons of the vertebrate hippocampus and the arthropod mushroom bodies, forebrain centers that support spatial navigation, allocentric memory, and associative learning (10,24).

Based on these findings, it has been postulated that the corresponding brain organisation in arthropods and vertebrates is an example of genealogical relationship that can be traced to a distant pre-Cambrian ancestor (8,24, 25). Indeed, evidence from soft tissue preservation in fossils of stem arthropods, suggests that the gross cerebral arrangements present in the four extant panarthropod lineages originated prior to the early Cambrian. This further implies that neural ground patterns attributed to the Panarthropoda may be both ancient and extremely stable over geological time (25). Here, ground patterns are defined as ancestral arrangements that are inherited with modification. However, due to the extreme rarity of such detailed fossil material, resolving correspondences across phyla has to rely instead on the identification of shared developmental rules and their outcomes (21,24). These correspondences are

105 expected to be defined by common gene regulatory (26) and character identity
106 networks (27) that convey positional information and identity to a species-specific
107 morphology, albeit often highly derived (28). Accordingly, it is the comparative analysis
108 of ground patterns across phylogenetic lineages, that allows the identification of
109 correspondences amongst cell types, tissue and organs and that informs about their
110 origins and genealogical relationships (29).
111 We applied this approach to compare the formation and function of the
112 *Drosophila* and vertebrate midbrain hindbrain boundary region. The vertebrate
113 Midbrain-Hindbrain Boundary (MHB) is positioned by adjacent *Otx* and *Gbx* activity
114 along the AP axis, and elaborated by region-specific expression of *Engrailed*, *Wnt*,
115 *Pax2/5/8*, and FGF8-mediated organizer activity (30-33). In *Drosophila*, the
116 corresponding boundary (henceforth referred to as the deutocerebral-tritocerebral
117 boundary, DTB) is defined by adjoining expression of *otd* and *unplugged (unpg),*
118 homologs of *Otx* and *Gbx*, respectively (34). The observation of these similar expression
119 patterns raises a number of questions: whether they reflect a shared developmental
120 program for the MHB and DTB; what adult brain structures derive from them; and
121 what their function might be. We hypothesized that if the DTB evolutionarily
122 corresponds to the vertebrate MHB, its formation would be mediated by gene
123 regulatory and character identity networks homologous to those driving MHB
124 formation. Furthermore, if true, we expected the DTB to give rise to circuits mediating
125 similar behaviours controlled by MHB-derived circuitry. Here we describe
126 experimental evidence verifying that the arthropod DTB indeed shares a ground
127 pattern organization with the vertebrate MHB, including correspondence of neural
128 circuits and their behavioral functions.
129

RESULTS

We focused on phylotypic (35) stage 11-14 embryos (**Fig. 1A, B**) to characterize morphological and molecular signatures of the developing *Drosophila* DTB. In addition to the adjoining *otd* and *unpg* expression and function reported earlier (34), we found specific domains of expression of the Pax2/5/8 homologs *shaven(sv)/dPax2* and *Pox neuro* (*Poxn*), as well as *engrailed* (*en*), *wingless* (*wg/dWnt*), *muscle specific homeobox* (*msh/dMsx*), *ventral nervous system defective* (*vnd/dNkx2*), and *empty spiracles* (*ems/dEmx*) (**Fig. 1C** and ***SI Appendix* Fig. S1A**). For axial patterning, we examined the expression and function of the key genes *otd* + *wg* (antero-posterior) and *msh* + *vnd* (dorsal-ventral), which revealed essential roles in DTB formation (***SI Appendix* Fig. S1B**).

A cardinal feature of the vertebrate MHB is its organizer activity, mainly mediated by the FGF8 effector molecule (30-33,36). Previous studies failed to identify a DTB-related function of the FGF8 homolog *branchless* and its receptor *breathless* in embryonic brain development of *Drosophila* (34). We now show that a second set of FGF8-like orthologs, *thisbe/FGF8-like1* and *pyramus/FGF8-like2* (37,38), and the FGF8 receptor heartless (*htl*), are expressed at the DTB (***SI Appendix* Fig. S2**). Consistent with earlier reports (37, 38), phylogenetic comparison of annotated protein sequences reveals that pyramus and thisbe, like branchless, are homologs of human and mouse FGF8, FGF17 and FGF18, as well as of the ascidian *Ciona intestinalis* FGF8/17/18, the Annelid *Capitella teleta* FGF and the Cnidarian *Nematostella vectensis* FGF8a (***SI Appendix* Fig. S3**). Similarly, phylogenetic comparison of annotated protein sequences reveals that the *Drosophila* FGF receptors breathless and heartless are homologs of human and mouse FGF receptors FGFR1-FGFR4, as well as of the ascidian *Ciona intestinalis* FGFR, the Annelid *Capitella teleta* FGFR and the Cnidarian *Nematostella*

155 *vectensis* FGFR (***SI Appendix* Fig. S4**). A functional role for FGF8 signalling at the
156 *Drosophila* DTB was revealed by altered *engrailed* expression patterns and
157 morphological defects in embryos homozygous for a deficiency, *Df(2R)BSC25*,
158 uncovering both *FGF8-like1* and *FGF8-like2* genomic loci, and of a *htl* null allele, $htl^{AB42-/-}$
159 (**Fig. 2A**). As indicated in the insets of panels a-c, these defects impact the prefrontal
160 commissure, (indicated by arrowheads), the size and integrity of longitudinal
161 connectives (indicated by the white brackets), as well as axonal projections, also
162 indicated by the white arrow in each panel at the bifurcation between prefrontal
163 commissure and longitudinal connective. These observations were further
164 substantiated by progressive changes and loss of the DTB expression patterns of *unpg*
165 and *ems* in *htl* null mutant embryos, between embryonic stages 12-16 (**Fig. 2B**, panels
166 d-f, compare to panels a-e). These data identify a role of FGF8-like signalling in the
167 maintenance of the embryonic DTB region in *Drosophila*. Given the expression pattern
168 of FGF8-like homologs in the brain and procephalic endoderm (***SI Appendix* Fig. S2**), it
169 remains to be shown whether cell autonomous or cell non-autonomous FGF8 signalling
170 is required for DTB formation and maintenance. Ectopic expression of the FGF8-like
171 homologue *thisbe* in *ems*-specific brain regions did not cause any detectable changes in
172 morphology or molecular signatures of the DTB region (***SI Appendix* Fig. S5**). Despite
173 conserved regulatory interactions between *otd/Otx* and *unpg/Gbx* (34), these data
174 indicate the absence of FGF8-mediated organizer activity in the embryonic DTB. We
175 conclude that direct or indirect FGF8-like signalling is required for the maintenance of
176 the embryonic DTB, but contrary to what is seen in vertebrate MHB development
177 appears not to organize the DTB region in *Drosophila*.
178 To identify the adult brain structures and associated functional modalities that
179 derive from the embryonic DTB, we utilized genetic tracing of neural lineages

employing the tracer line *UAS-mCD8::GFP, tub-FRT-CD2-FRT-Gal4, UAS-FLP/CyO GMR Dfd YFP* (39, 40). We first deployed *en-Gal4* as driver, which combined with the tracer line causes permanent GFP labelling of progeny of *engrailed*-expressing cells, even when those progeny themselves no longer express *engrailed*. This is due to the fact that the genetic tracer line carries an FLP-out cassette, in which a strong and ubiquitous *tubulin* enhancer is separated from a Gal4 transcriptional activator by CD2, flanked by a pair of FRT sites, thereby preventing Gal4 expression. As soon as the genetic tracer line is crossed with an independent Gal4 driver, in this case *en-Gal4*, it initiates UAS-FLP expression as part of the genetic tracer, which in turn induces recombination at the CD2-flanking FRT sites. As a result, the CD2 element is excised, resulting in a fusion of the *tubulin* enhancer to drive Gal4 which in turn leads to expression of *UAS-mCD8::GFP* in that cell and all of its progeny. Since *tubulin* is constitutively active, the genetic tracing thus permanently marks targeted cells and their progeny, thereby revealing lineage-relationship even if some of the traced cells no longer express *engrailed*.

Genetic tracing of *engrailed* expressing lineages of the embryonic neuroectodermal DTB (***SI Appendix* Fig. S6A-C**) identified neurons and projections of the antennal mechanosensory and motor centre (AMMC) and select antennal glomeruli in the adult brain (**Fig. 1D, E** and ***SI Appendix* Fig. S6D-F**). We also determined the fate of Poxn expressing cells, which in the embryonic brain are located next to DTB-specific engrailed lineages (**Fig. 3A-H**). In the adult brain, Poxn expressing cells are present in a caudal volume of the deutocerebrum called the Wedge receiving auditory and gravity sensing afferents (41), where its interneurons, as do En-positive cells, express the neurotransmitter GABA (**Fig. 3I-P**). Genetic tracing revealed that the Poxn-expressing AMMC/Wedge neurons derive from DTB lineages. Furthermore, their region-specific projection patterns resemble the mechanosensory pathway architecture of local

205 interneurons and projection neurons (**Fig. 3Q-T**) previously identified for the AMMC
206 and the Wedge (41-44).
207 The *Drosophila* AMMC and Wedge neuropils comprise neurons that mediate
208 auditory, vestibular, mechanosensory and somatosensory information processing in
209 pathways with similarities to the mammalian auditory and vestibular pathways (41-
210 44). In vertebrates, auditory, vestibular, somatosensory and motor information, are
211 processed by neural populations of the tectum and cerebellum, adult brain structures
212 derived from the MHB region (36, 45). These fate-mapping studies in mice did not
213 identify specific neural circuits within the tectum and cerebellum, which limits a
214 structural comparison until more refined studies at cell and circuit resolution in
215 vertebrates become available.
216 Despite these limitations, we reflected on whether functional comparisons
217 might be possible. The tectum and cerebellum receive auditory and vestibular, as well
218 as motor information and, among other functions, are important for balance, body
219 posture, sensorimotor integration and motor coordination (1,4,46). To test whether
220 DTB-derived circuits in *Drosophila* might exert similar functions, the GAL4-UAS system
221 was used to express *Tetanus-Toxin-Light-Chain* (*TNT*) and inhibit synaptic transmission
222 (47) in subsets of AMMC neurons (42). Flies were tested for their startle-induced
223 negative geotaxis (SING) response, which after being shaken to the bottom in a test
224 tube quantifies their ability to right themselves and climb upwards (48). Except
225 *R19E09>TNT*, all of the tested genotypes showed significantly impaired SING behaviour
226 (**Fig. 4A** and ***SI Appendix* Table S1**). This includes R79D08, R45D07 and R30A07 that
227 target AMMC neurons and co-express *engrailed*, or encompass *Poxn* and *engrailed*
228 expression domains (**Fig. 4C-I**). Of note, several of the tested genotypes showed

difficulties with balance and to right themselves, as exemplified for *R52F05>TNT* compared to control (***SI Appendix* Movies S1** and **S2**).

To further analyze AMMC-mediated motor coordination, we employed video-assisted motion tracking and recorded freely moving flies (**Fig. 4B**). During 135-minute recordings, activity bouts and movement trajectories were analyzed to quantify locomotion parameters: frequency of episodic movements, how often they were initiated, their length and average velocity, as well as the duration and frequency of pauses. Response to sensory stimulation triggered by repeated mechanical shock was also recorded. *UAS-TNT* expression by *R79D08* (*CG9650/dZic-B*)-Gal4, which targets zone B (41-43) of the AMMC (**Fig. 4C**, arrows), significantly impaired overall activity and duration, with fewer actions initiated, shorter episodes of activity and their intervals, reduced velocity and distances travelled (**Fig. 4D** and ***SI Appendix* Fig. S7A**). *UAS-TNT* expression by *R88B12* (*en/inv*)-Gal4 targeting zone A (41-43) of the AMMC (**Fig. 4E**, arrows) significantly impaired average and pre-stimulus speed, with shorter bouts of activity, together resulting in less distance travelled (**Fig. 4F** and ***SI Appendix* Fig. S7B**). Comparable motor phenotypes were seen with *R30A07* (*NetA*)-Gal4, which targets AMMC neurons that co-express *engrailed* (**Fig. 4G**, arrowheads; **Fig. 4H** and **Fig. S7C**), but not with *R45D07* (*hth*)-Gal4 targeting parts of the AMMC-specific giant fiber system (**Fig. 4I, J** and ***SI Appendix* Fig. S7D**). Together, with SING data, our behavioural observations establish essential functions of the AMMC for sensorimotor integration (41,42,44), balance, righting reflex and motor coordination in *Drosophila,* behavioural manifestations similar to MHB region-derived circuits in vertebrates.

Our findings thus far establish correspondences between *Drosophila* DTB and vertebrate MHB at multiple levels including adult brain circuits and the behaviours they regulate. We hypothesized that this will be reflected in commonalities among

character identity networks of DTB and MHB that are mediated by homologous gene regulatory networks (27,28). To test this hypothesis, we screened the Janelia Gal4 collection (49) for cis-regulatory elements (CREs) mediating the spatiotemporal expression of developmental genes controlling DTB formation in *Drosophila.* We identified CREs for *msh*, *vnd*, *ems*, and for *thisbe/FGF8-like1* (***SI Appendix* Fig. S8A-E)**, genes that are essential for the formation and/or maintenance of the embryonic DTB (**Fig. 2** and ***SI Appendix* Fig. S1** and ref. 14). In addition, we identified CREs for *Wnt10*, *Sex combs reduced/Hox5;* the *Drosophila* homologs of *zinc finger of the cerebellum (ZIC), odd-paired* (*opa/dZic-A*) and *CG9650/dZic-B;* of *Purkinje cell protein 4* (*PCP4*), *igloo* (*igl/dPCP4*); of *Ptf1a, Fer1/dPtf1a*, and of *Lim1* (***SI Appendix* Fig. S8F-J**). All mammalian homologs of these genes have been implicated in vertebrate MHB formation and the specification of midbrain-cerebellar circuitry (30-34, 50) (see ***SI Appendix* Table S2**).

We also identified CREs for *dachshund* (*dac*) and the *Pax2* homologue *shaven* (*sv/dPax2*). *dac*-specific CRE *R65A11-Gal4* targeted *UAS-mCD8::GFP* expression to the DTB region, in derived lineages of the embryonic brain and to the AMMC in a pattern encompassed by DTB-specific *engrailed* and *Poxn* expression domains (**Fig. 5A-C**). The regulatory element VT51937 (51) located in an intron of *sv/dPax2* targets Gal4 expression in a segment-specific pattern similar to endogenous *sv/dPax2*, including DTB expression domains (***SI Appendix* Fig. S9A-F**). *VT51937-Gal4* mediated genetic tracing also identified cells and projections in the AMMC (***SI Appendix* Fig. S9G**). Together these data identify CREs of the DTB-AMMC character identity network in *Drosophila* that mediate the spatiotemporal expression patterns of genes that are homologous to genes involved in the formation and specification of the vertebrate MHB and derived midbrain-cerebellar circuitry (30-34,50).

Given the correspondences between the *Drosophila* DTB and vertebrate MHB gene regulatory and character identity networks, we asked whether the cis-regulatory elements (CREs) that control the expression of these genes might be conserved. To identify conserved cross-phylum CREs, we utilized DTB-AMMC specific regulatory sequences and applied bioinformatics tools (52) including VISTA (53), MLAGAN (54) and EMBOSS MATCHER (55) to screen for corresponding CREs in mouse and human genomes (56). Stringent selection criteria (57) were applied to identify CRE sequences *(i)* that are linked to homologous genes in the different species; *(ii)* with a minimum of 62% sequence identity over at least 55 base pairs and at least $1e^{-1}$ confidence level as the BLAST e-value; *(iii)* that are not un-annotated protein sequences; and *(iv)* that are not repetitive elements. Additionally, BLAT (58) was applied for searching vertebrate genomes using the same selection criteria.

We first analyzed the DTB-AMMC-specific CRE of *sv/dPax2* (= VT51937 sequence) and identified non-coding CREs for mouse *Pax2* and human *PAX2* with extensive sequence similarities (***SI Appendix* Fig. S9H** and **data set S1**), and comparable intragenic location (***SI Appendix* Fig. S9I**). To validate the significance of these sequences, we carried out BLAST/BLAT searches of the *sv/PAX2* conserved sequence against the *Drosophila melanogaster*, *Mus musculus* and human genome sequences. These data revealed that in *Drosophila melanogaster*, only the *sv/dPax2*-related CRE of 255 basepairs with 1.6E-142 Blast e-value matches the cutoff criteria of >62% sequence identity over at least 55 basepairs with minimum 1^{e-1} confidence level as the BLAST e-value, whereas all other identified sequences were of 18bp or shorter length (***SI Appendix*** data set S1). Of note, BLAT searches of the *sv/PAX2* conserved sequence against the mouse and human genome identified CREs only related to mouse *Pax2* and human *PAX2* and no other genomic sequence (***SI Appendix*** data set S1). In

303 addition, using the JASPAR algorithm (59), we identified potential transcription factor
304 binding sites that match stretches of the MLAGAN-aligned *sv/PAX2* conserved CRE
305 sequence (***SI Appendix*** data set S1).
306 Following this strategy, we used DTB-AMMC-specific CRE sequences for
307 *dachshund* and *engrailed/invected* and identified human CREs conserved among
308 vertebrates that direct MHB-specific expression in mouse for the *dachshund* homologs
309 *Dach1/DACH1* (**Fig. 5D-F**) and for the *engrailed/invected* homologs *Engrailed2/EN2* (***SI***
310 ***Appendix* Fig. S10**). Further bioinformatics analysis identified core CRE sequences
311 associated with *dac* and *DACH1*, *en/inv* and *EN2* and *sv* and *PAX2* in *Drosophilidae* and
312 vertebrate genomes (***SI Appendix* data sets S1-S3**). As was the case for the *sv/PAX2*
313 conserved CRE sequences, BLAST/BLAT searches against the *Drosophila melanogaster*,
314 *Mus musculus* and human genomes revealed for the *Drosophila melanogaster* genome
315 that only the *en/inv*-related CRE of 616 basepairs with 0.0 Blast e-value (***SI Appendix***
316 data set S2) and *dac*-related CRE of 247 basepairs with 9.4E-138 Blast e-value (***SI***
317 ***Appendix*** data set S3) match the cutoff criteria, whereas all other identified sequences
318 are of maximum 23bp (for *en/inv*-related) or 30bp (for *dac*-related) or shorter length.
319 Also, BLAT searches of the *invected/engrailed-EN2* and *dachshund/DACH1* conserved
320 sequences against the mouse and human genome only identified CREs for mouse *En2*
321 and human *EN2*, and mouse *Dach1* and human *DACH1* respectively, whereas no other
322 genomic sequences can be identified (***SI Appendix*** data set S2 and 3). Furthermore, and
323 again using the JASPAR algorithm, we identified potential transcription factor binding
324 sites that match stretches of the MLAGAN-aligned *invected/engrailed-EN2* (***SI Appendix***
325 data set S2) and *dachshund/DACH1* (***SI Appendix*** data set S3) conserved CRE
326 sequences, respectively. Together these data suggest ancestral non-coding regulatory

sequences and their function predate the radiation of insect-specific DTB and vertebrate-specific MHB circuits and morphologies.

DISCUSSION

We have identified gene regulatory and character-identity networks that underlie the formation of the deutocerebral-tritocerebral boundary in *Drosophila.* Mutant analyses reveal that *otd* + *wg* and *msh* + *vnd*, acting along the AP and DV body axes, respectively, are required for the formation of the embryonic DTB, and that FGF8-like signaling is necessary for its developmental maintenance. Genetic tracing experiments, together with the analysis of cis-regulatory elements for *engrailed/invected*, *dachshund* and *shaven/dPax2*, as well as behavioral analysis after synaptic inactivation show that embryonic DTB lineages give rise to neural circuits in the AMMC/Wedge complex of the adult brain that mediate balance and motor coordination in *Drosophila*. Together these findings establish a ground pattern of DTB formation and derived circuit function in *Drosophila* that corresponds to the ground pattern, gene regulatory and character identity networks of the vertebrate MHB and derived midbrain-cerebellar circuits (***SI Appendix* Fig. S11**).

In contrast to a previous gene expression study pointing towards the protocerebral/deutocerebral boundary (60), our findings based on gene regulation and function identify the *Drosophila* DTB as the boundary between the rostral brain and its genetically distinct caudal nervous system. These data imply that caudal domains of the arthropod deutocerebrum and its circuits in *Drosophila* correspond to the vertebrate MHB and its derived principle proprioceptive circuits (see ***SI Appendix* Fig. S11**). It must be emphasized here that these circuits are not ascribed to the cerebellum, the anlage of which forms as an asegmental volume within *Gbx2* and non-*Hox* expression

352 domains of the developing MHB (30-33,36,61). Rather, our comparative analysis
353 suggests that vestibular/balance-related circuits characterize the ancestral territory
354 from which the DTB and MHB evolved, and that MHB-derived cerebellar circuitry is a
355 vertebrate innovation. In support of this notion, FGF8 signalling in vertebrates acts as a
356 secondary organizer in boundary development of the MHB and by promoting growth
357 essential for the formation of tectum and cerebellum (30-34,36,61). We did not observe
358 FGF8-like organizer activity in flies but a role in the maintenance of the DTB boundary,
359 suggesting an ancestral boundary-related function for FGF8 (61). A comparable
360 phenotype has been described in ascidian larvae mutant for *Ffg8/17/18* (62). Normally
361 expressed in the visceral ganglion, knockdown of *Fgf8/17/18* altered *Otx*, *Engrailed*,
362 and *Pax2/5/8* gene expression in the anterior adjacent neck region of *Ciona*, essentially
363 leading to a posterior expansion of rostral CNS identity (62). These data suggest that
364 similar to the situation in the embryonic brain of *Drosophila*, FGF8 signalling in the
365 tadpole larvae of *Ciona intestinalis* delineates the boundary between the rostral brain
366 and the caudal nervous system. Moreover, the observed absence of extended
367 proliferative activity in *Drosophila* (34 and this paper) and *Ciona* (62) suggests that
368 growth-related organizer activity of FGF8 is a vertebrate innovation, whereas the
369 boundary region is defined by expression patterns of genes homologous to those
370 observed at the DTB/MHB (**Fig. 6**). Indeed, in spite of comparable expression domains
371 (11,12,63,64), no phenotypic cerebellum is found in ascidians, nor in hemichordates
372 and cephalochordates; none of which can be assumed as proxies for ancestral
373 vertebrates, but all of which may represent highly derived and evolutionary simplified
374 crown species. However, ancestral circuits mediating vestibular and motor (balance)
375 coordination, which are specified by genes and regulatory networks homologous to
376 those described in this paper, are found in lampreys and hagfish, the persisting ancient

lineages of early vertebrates (2), and maybe also in the chordate brain of *Ciona intestinalis* tadpole larvae (65).

The observed correspondences in circuit formation extend to behaviours they regulate. Synaptic inactivation of DTB-derived sub-circuits of the AMMC/Wedge complex, results in flies with impaired balance, defective action initiation and maintenance, and compromised sequences of motor actions. These AMMC circuits have been shown to mediate auditory and vestibular information processing and coordination (41,42), suggesting that the DTB-derived AMMC/Wedge circuits integrate mechanosensory submodalities for motor homeostasis. These functions correspond to activity of MHB-derived acoustic and vestibular receptor pathways in vertebrates (41,44) which, when impaired in inherited disorders, affect both auditory and vestibular functions such as seen in ataxic patients (66). The observed correspondences therefore suggest that similar to MHB-derived circuits, the DTB-derived AMMC/Wedge circuits in arthropods are required for sensorimotor integration, body posture and motor coordination. These findings identify correspondences between ground patterns of the insect DTB and vertebrate MHB that extend beyond spatiotemporal expression patterns and functions of homologous genes, to neural circuits and behavior.

The present results identify CREs associated with highly conserved developmental control genes (67-70) regulating boundary formation between the rostral brain and the caudal nervous system in insects and vertebrates. Core elements of the identified CREs are highly conserved as demonstrated by sequence identities across large phylogenetic distances and by the identification of potential transcription factor binding sites that are located within these conserved CREs. Although there is evidence that orthologous CREs can be completely divergent at the sequence level (62),

our findings demonstrate that the identified conserved CREs are employed for the formation of circuits with comparable roles in neuronal processing. These data suggest the identified CREs are ancestral non-coding regulatory sequences with which insect-specific DTB and vertebrate-specific MHB circuits and morphologies evolved.

In conclusion, the corresponding ground patterns of insect DTB and vertebrate MHB suggest the early appearance in bilaterian evolution of a cephalic nervous system that evolved predictive motor homeostasis before the divergence of the protostome lineages and before the origin of deuterostomes. Based on the observed correspondences, we hypothesize that the retention across phyla of conserved regulatory mechanisms is necessary and sufficient for the formation and function of neural networks for adaptive behaviors common to all animals that possess a brain.

MATERIALS AND METHODS

Fly Strains and Genetics. The wild-type strain used was Oregon R. The following mutant alleles and characterization constructs were used to investigate expression and function: *P{en2.4-Gal4}e16E, UAS-mCD8::GFP^{LL5} and poxn^{brain}-Gal4* as well as *UAS-mCD8::GFP, tub-FRT-CD2-FRT-Gal4, UAS-FLP/CyO GMR Dfd YFP* (71); *otd^{JA101}* (14); *P{lacZ}unpg^{f85}* (an *unpg-lacZ* reporter gene that expresses cytoplasmic β-galactosidase in the same pattern as endogenous *unpg*) (34); *P{lacZ}Pax2$^{\Delta 122}$* (a *Pax2-lacZ* reporter gene that expresses β-galactosidase in the same pattern as endogenous *Pax2*) (34); *P{3′lacZ}unpg^{r37}* (*unpg* null allele with a *unpg-lacZ* reporter gene that expresses nuclear β -galactosidase in the same pattern as endogenous *unpg*) (34); *wg^{CX4}* and *wg-lacZ* (Bloomington); *msh$^{\Delta 68}$* (72)*; vnd^{6}* (73)*;* the deficiency *Df (2R) BSC25* that removes the FGF8-like 1 and FGF8-like 2 loci together with adjacent regions and *htl^{AB42}* (37) *UAS-FGF8-like 1* (37); and *ems2.6 (72.5)-Gal4* (73).

To generate *Poxn^{brain}-Gal4* flies, the *Poxn* brain enhancer (74) was amplified by PCR from genomic DNA. The PCR product was subcloned into *pPTGal* vector using *XbaI* and *NotI* sites, followed by sequencing; the genomic region 2R:11723830 to 11725559 was inserted into *pPTGal*. Primer sequences are:

forward, *5'-gctcattaatgaccatgaaa-3'*;

reverse, *5'-aagcggccgcgttaagtaacgctcggtgg-3'*.

Transgenesis was performed by BestGene Inc (CA, USA).

For lineage tracing, the following strains were used: *w^{1118}* (control), *P{en2.4-Gal4}e16E, UAS-mCD8::GFP^{LL5}*, or *poxn^{brain}-Gal4* were crossed to *UAS-mCD8::GFP, tub-FRT-CD2-FRT-Gal4, UAS-FLP/CyO GMR Dfd YFP*. Offspring were raised at 18°C to suppress random leaky FLP activity.

For behavioural experiments, we used *UAS-TNT-E* (47) crossed to AMMC-specific *Gal4* lines. Corresponding controls for *Gal4* driver and for UAS responder line, were generated by backcrossing to *w^{1118}*. All behavioural experiments were carried out in a temperature-controlled chamber at 25°C.

In situ Hybridization, Immunocytochemistry and Image Analysis. For in situ hybridization experiments, digoxigenin-labelled sense and antisense RNA probes were generated in vitro with a DIG labelling kit (Roche diagnostics) and hybridized to *Drosophila* whole-mount embryos, following standard procedures (75).

Whole-mount immunocytochemistry was performed as previously described (76,77). Primary antibodies were rabbit anti-Otd (34), used 1:100; rabbit anti-Msh (72), used 1:500; rabbit anti-Vnd (73), used 1:200; rabbit anti-sv/dPax2 (34), used 1:50; monoclonal anti-Poxn antibodies (78), used 1:20; rabbit anti-HRP (FITC-conjugated, Jackson Immunoresearch), used 1:50; mouse anti-En (Developmental

Studies Hybridoma Bank, DSHB), used 1:1; rabbit anti-β-Gal, used 1:200 (Milan analytica); mouse anti-β-Gal (DSHB), used 1:100; rabbit anti-Lab (79), used 1:50; rat anti-Ems (80), used 1:2000; mouse anti-Bruchpilot (nc82, DSHB), used 1:20; mouse anti-Synapsin (3C11, DSHB), used 1:50; rabbit anti-GFP (Invitrogen), used 1:500; rabbit anti-GABA (Sigma-Aldrich), used 1:1000. Secondary antibodies were Alexa-568-conjugated goat anti-mouse, Alexa-568-conjugated goat anti-rabbit, Alexa-568-conjugated goat anti-rat, Alexa-488-conjugated goat anti-mouse, Alexa-488-conjugated goat anti-rabbit, and Alexa-488-conjugated goat anti-rat (Molecular probes), all used 1:150. Embryos, larval CNS and adult brains were mounted in Vectashield H-1000 (Vector).

Fluorescence samples were scanned and recorded with a Leica TCS SP5 confocal microscope. Z-projections were created and analysed using FIJI. Images were processed using Adobe Photoshop and figures constructed in Adobe Illustrator.

Startle Induced Negative Geotaxis Assay (SING). Groups of 10 flies with shortened wings of the same age, sex and genotype were placed in a vertical column (19 cm long, 2cm diameter). The wings were clipped under sedation (with CO_2) at least 24 hours prior to testing. They were suddenly startled by gently tapping them down, to which *Drosophila* responds by climbing up. After 10 seconds, it was counted how many flies were above the 2cm mark and the trial was repeated 15 times for each tube and the average was calculated. For each genotype 10 groups of females and 10 groups of males were tested. Flies were reared at 25°C and were maintained under 12hr light/dark cycle. Flies with an average age of 5 days were tested at RT, under the same light conditions. All assays were performed at the same time of day.

Motor Behavior Analysis. Control and experimental flies were reared at 18°C and adult mated females up to 5 days post-eclosion were transferred to 25°C for behavioural analyses. Mechanical stimuli trains consisted of 5 pulses of 200ms each, spaced by 800ms. Motor behaviour parameters were determined as previously described (38,81,82).

Bioinformatics Analyses and Identification of Cis-Regulatory Elements (CREs). The Janelia Gal4 collection (http://flweb.janelia.org/cgi-bin/flew.cgi) (49) was screened for AMMC-specific GFP expression patterns. All hits were cross checked to be verified/excluded from the Janelia/Bloomington list (see https://bdsc.indiana.edu/stocks/gal4/gal4_janelia_info.html). For each hit, annotated left and right primers were used to BLAST the *Drosophila* genome annotated at the *Ensembl* genome browser (http://www.ensembl.org/index.html) to determine the position and sequence within genome version BDGP6, or where known the sequence was extracted from the respective gene map annotation in JBrowse (http://flybase.org/). The resulting sequence was compared against available VT enhancer sequences and their annotated expression pattern determined for DTB expression patterns (http://enhancers.starklab.org/) (51). *Drosophila*-specific, non-coding regulatory sequences were then used for BLAT searches (58) to screen the mouse (GRCm38.p5) and human (GRCh38.p7) genome annotated at Ensemble (https://www.ensembl.org/index.html) to identify any potential corresponding sequences. Any matching sequences were scrutinized for further analysis on the basis of criteria that have been used previously to identify transphyletic cis-regulatory DNA sequences (57). These criteria were:

(i) the sequences are linked to the same homologous genes in the different species;

(ii) there is a minimum of 62% sequence identity over at least 55 bp with minimum 1e^{-1} confidence level as the BLAST e-value;

(iii) the CREs are not un-annotated protein sequences; and

(iv) the CREs are not repetitive elements.

Because the algorithm used by BLAST/BLAT does not support the first stated threshold criteria, our search was restricted to and thus focused on the genomic regions encompassing homologous genes. In the case of more than one homolog (e.g. *DACH1* and *DACH2*) both homologues were included in the search. In the case of *EN2*, we additionally included the intergenic region between the two *Drosophila* homologues, *invected* and *engrailed*.

The resulting sequences were then used for refined comparisons using pair-wise and multiple sequence alignment algorithms including EMBOSS Matcher and t-coffee (http://www.ebi.ac.uk/services). To carry out sequence alignments, which automatically indicated whether any CNS regulatory elements might be covered by the input sequences, we used the MLAGAN algorithm (http://genome.lbl.gov/vista/lagan/submit.shtml). Detected CNS CREs were then further scrutinized using the VISTA enhancer browser (https://enhancer.lbl.gov/frnt_page_n.shtml), which provides human and mouse regulatory sequences and their expression pattern at embryonic stage E11.5 in transgenic mouse embryos expressing LacZ under the control of the respective regulatory sequence (53,56). The relevant images of LacZ expression were extracted and reproduced with permission by Dr. Len Pennacchio, Lawrence Berkeley National Laboratory. Identified MHB-specific regulatory sequences were utilized to perform multiple sequence alignment with the respective DTB->AMMC-specific regulatory elements; any matches were re-confirmed and quantified using the EMBOSS Matcher

and CLUSTAL Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) algorithms.

Potential transcription factor binding sites (TFBS) that match stretches of the MLAGAN-aligned conserved CRE sequences were identified using the JASPAR algorithm ((http://jaspar.genereg.net/) (59).

Maximum likelihood phylogenetic trees were inferred by the Genome-to-Genome-Distance Calculator (GGDC) available at the Leibniz Institute DSMZ (https://ggdc.dsmz.de/phylogeny-service.php) from a MUSCLE multiple sequence alignment with Randomized Axelerated Maximum Likelihood (83).

Statistical Analysis. Each data set was tested for normality using the Anderson-Darling test with $\alpha = 0.05$. If every data set under comparison was normal and the variances were similar (Hartley's fmax was calculated in each case and used as a cut-off for variance ratio), then a one-way ANOVA test was used to determine whether any differences existed between groups. If significance was found for ANOVA with $\alpha = 0.05$, then pair-wise comparisons were made using a post hoc Tukey-Kramer test, again with $\alpha = 0.05$. If any of the data sets was found not to be normally distributed, then a Kruskal-Wallis test was used to determine any overall differences between the groups with $\alpha = 0.05$. If significance was achieved, a post hoc pairwise Mann-Whitney U test with Dunn-Sidak correction was used to compare groups with $\alpha = 0.05$. For each test group, two controls were used corresponding to the two genetic elements that were altered in the group under analysis. For example, *R11A07>TNT* would be tested against TNT control and *R11A07>w1118*. For a result to be considered significant the experimental group had to be significantly different from both controls and the controls not be significantly different from one another. All calculations were performed using

MATLAB.

ACKNOWLEDGEMENTS. We thank H. Karten, C. Ragsdale, A. Kamikouchi, A. Delogu, M. Fanto, and T.J. Weidemann for discussions; G.M. Rubin and L. Pennacchio for CRE data; D. Diaper and M. Dyson for graphics; and U. Walldorf, M. Noll, S. Carroll, A. Mueller, M. Michaelson, M. Landgraf, A. Ghysen, A. Gould, as well as the Developmental Studies Hybridoma Bank and the VDRC and Bloomington Stock Centres for providing antibodies and fly strains. This work was supported by a PhD fellowship from CAPES Foundation – Ministry of Education of Brazil to J.C.B. (BEX 13162/13-6); an IoPPN-King's Independent Researcher Award to B.K.; an FWO post-doc grant to L.V.B.; the Flemish FWO (G065408.N10 and G078914N) to P.C.; the National Science Foundation under Grant No. 1754798 awarded to N.J.S.; and the UK Medical Research Council (G0701498; MR/L010666/1), the UK Biotechnology and Biological Sciences Research Council (BB/N001230/1) and the MND Association (Hirth/Nov15/914-793; Hirth/Oct13/6202; Hirth/Mar12/6085; Hirth/Oct07/6233) to F.H.

AUTHOR CONTRIBUTIONS: J.C.B., B.H., Z.N.L., L.V.B. and B.K. performed experiments; F.H. and M.G. performed CRE analysis and phylogenetic comparisons; J.D. performed statistical tests for SING data. All authors analysed the data; F.H., P.C. and N.J.S. prepared the manuscript; F.H. conceived and designed the project.

COMPETING INTEREST: B.K. is co-founder of BFKLab LTD. The remaining authors declare no competing financial interest.

ADDITIONAL INFORMATION: Supplementary information includes supplementary Figs. S1 to S11, Tables S1 to S2; captions for movies S1 to S2; captions for databases S1 to S3 and references for SI reference citations.

Data Availability: All data of this study are included in the manuscript and SI Appendix.

REFERENCES

1. Robertson, B., Kardamakis, A., Capantini, L., Pérez-Fernández, J., Suryanarayana, S.M., Wallén, P., Stephenson-Jones, M., Grillner, S. The lamprey blueprint of the mammalian nervous system. *Prog. Brain Res.* **212**, 337-49 (2014).
2. Sugahara, F., Pascual-Anaya, J., Oisi, Y., Kuraku, S., Aota, S., Adachi, N., Takagi, W., Hirai, T., Sato, N., Murakami, Y., Kuratani, S. Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain. *Nature* **531**, 97-100 (2016).
3. Brown, M., Keynes, R., Lumsden, A. *The Developing Brain*. (Oxford Univ. Press, 2001).
4. Strausfeld, N.J. *Arthropod Brains: Evolution, Functional Elegance and Historical Significance.* (Belknap Press of Harvard Univ. Press, 2012).
5. Cong, P., Ma, X., Hou, X., Edgecombe, G.D., Strausfeld, N.J. Brain structure resolves the segmental affinity of anomalocaridid appendages. *Nature* **513**, 538-542.
6. Ghysen, A. The origin and evolution of the nervous system. *Int. J. Dev. Biol.* **47**, 555-562 (2003).
7. Hirth, F., Reichert, H. "Basic nervous system types: one or many?" in *Evolution of Nervous Systems*, T.H. Bullock, L.A. Krubitzer, T.M. Preuss, J.L.R. Rubenstein, N.J. Strausfeld, G.F. Striedter GF, Eds. (Elsevier, 2005), Vol I, pp 55-72.
8. Hirth, F. On the origin and evolution of the tripartite brain. *Brain Behav. Evol.* **76**, 3-10 (2010).
9. Strausfeld, N.J., Hirth, F. Deep homology of arthropod central complex and vertebrate basal ganglia. *Science* **340**, 157-161 (2013).

10. Wolff, G.H., Strausfeld, N.J. Genealogical correspondence of a forebrain centre implies an executive brain in the protostome-deuterostome bilaterian ancestor. *Philos. Trans. R. Soc. Lond. B Biol Sci.* **371**, 20150055 (2016).
11. Holland, N.D. Early central nervous system evolution: an era of skin brains? *Nat. Rev. Neurosci.* **4**, 617-627 (2003).
12. Holland, L.Z. *et al.* Evolution of bilaterian nervous systems: a single origin? *EvoDevo* **4**, 27 (2013).
13. Hirth, F., Reichert, H. Conserved genetic programs in insect and mammalian brain development. *Bioessays* **21**, 677-684 (1999).
14. Hirth, F. *et al.* Developmental defects in brain segmentation caused by mutations of the homeobox genes *orthodenticle* and *empty spiracles* in *Drosophila*. *Neuron* **15**, 769-778 (1995).
15. Simeone, A. Otx1 and Otx2 in the development and evolution of the mammalian brain. *EMBO J.* **17**, 6790-6798 (1998).
16. Nagao, T. *et al.* Developmental rescue of Drosophila cephalic defects by the human Otx genes. *Proc. Natl. Acad. Sci. USA* **95**, 3737-3742 (1998).
17. Leuzinger, S. *et al.* Equivalence of the fly orthodenticle gene and the human OTX genes in embryonic brain development of Drosophila. *Development* **125**, 1703-1710 (1998).
18. Acampora, D. *et al.* Murine Otx1 and Drosophila otd genes share conserved genetic functions required in invertebrate and vertebrate brain development. *Development* **125**, 1691-1702 (1998).
19. Acampora, D. *et al.* OTD/OTX2 functional equivalence depends on 5' and 3' UTR-mediated control of Otx2 mRNA for nucleo-cytoplasmic export and epiblast-restricted translation. *Development* **128**, 4801-4813 (2001).
20. Hanks, M.C. *et al. Drosophila* engrailed can substitute for mouse Engrailed1 function in mid-hindbrain, but not limb development. *Development* **125**, 4521-4530 (1998).
21. Strausfeld, N.J., Hirth, F. Homology versus convergence in resolving transphyletic correspondences of brain organization. *Brain Behav. Evol.* **82**, 215-219 (2013).
22. Fiore, V.G., Dolan, R.J., Strausfeld, N.J., Hirth, F. Evolutionarily conserved mechanisms for the selection and maintenance of behavioural activity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20150053 (2015).

638 23. Fiore, V.G. *et al.* Changing pattern in the basal ganglia: motor switching under
639 reduced dopaminergic drive. *Sci. Rep.* **6**, 23327 (2016).
640 24. Strausfeld, N.J. & Hirth, F. Introduction to 'Homology and convergence in nervous
641 system evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150034 (2016).
642 25. Strausfeld, N.J., Ma, X., Edgecombe, G.D. Fossils and the evolution of the arthropod
643 brain. *Curr. Biol.* **26**, R989–1000 (2016).
644 26. Erwin, D.H., Davidson, E.H. The evolution of hierarchical gene regulatory
645 networks. *Nat. Rev. Genet.* **10**, 141-148 (2009).
646 27. Wagner, G.P. The developmental genetics of homology. *Nat. Rev. Genet.* **8**, 473-
647 479 (2007).
648 28. Wagner, G.P. *Homology, Genes and Evolutionary Innovation.* (Princeton Univ.
649 Press, 2014).
650 29. Arendt, D. *et al.* The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744-757
651 (2016).
652 30. Rhinn, M. & Brand, M. The midbrain-hindbrain boundary organizer. *Curr. Opin.*
653 *Neurobiol.* **11**, 34-42 (2001).
654 31. Wurst, W. & Bally-Cuif, L. Neural plate patterning: upstream and downstream of
655 the isthmic organizer. *Nat. Rev. Neurosci.* **2**, 99-108 (2001).
656 32. Dworkin, S. & Jane, S.M. Novel mechanisms that pattern and shape the midbrain
657 hindbrain boundary. *Cell. Mol. Life Sci.* **70**, 3365-3374 (2013).
658 33. Harada, H., Sato, T. & Nakamura, H. Fgf8 signaling for development of the
659 midbrain and hindbrain. *Dev. Growth Differ.* **58**, 437-445 (2016).
660 34. Hirth, F. *et al.* An urbilaterian origin of the tripartite brain: developmental genetic
661 insights from *Drosophila*. *Development* **130**, 2365-2373 (2003).
662 35. Levin, M. *et al.* The mid-developmental transition and the evolution of animal
663 body plans. *Nature* **531**, 637–641 (2016).
664 36. Sato, T. & Joyner, A.L. The duration of FGF8 isthmic organizer expression is key to
665 patterning different tectal-isthmo-cerebellum structures. *Development* **136**, 3617-
666 3626 (2009).
667 37. Gryzik, T. & Mueller, H.-A. J. FGF8-like1 and FGF8-like2 encode putative ligands of
668 the FGF receptor Htl and are required for mesoderm migration in the *Drosophila*
669 gastrula. *Curr. Biol.* **14**, 659-667 (2004).

38. Stathopoulos, A., Tam., B., Ronshaugen, M., Frasch, M. & Levine, M. pyramus and thisbe: FGF genes that pattern the mesoderm of Drosophila embryos. *Genes Dev.* **18**, 687-699 (2004).
39. Shaw, R.E. *et al. In vivo* expansion of functionally integrated GABAergic interneurons by targeted increase in neural progenitors. *EMBO J.* e98163 (2018).
40. Bridi, J.C., Ludlow, Z.N. & Hirth, F. Lineage-specific determination of ring neuron circuitry in the central complex of *Drosophila*. *Biol. Open* **8**, bio045062 (2019).
41. Kamikouchi, A. *et al.* The neural basis of *Drosophila* gravity-sensing and hearing. *Nature* **458**, 165-171 (2009).
42. Vaughan, A.G., Zhou, C., Manoli, D.S. & Baker, B.S. Neural pathways for the detection and discrimination of conspecific song of *D. melanogaster*. *Curr. Biol.* **24**, 1039-1049 (2014).
43. Matsuo, E. *et al.* Organization of projection neurons and local neurons of the primary auditory center in the fruit fly *Drosophila melanogaster*. *J. Comp. Neurol.* **524**, 1099-1164 (2016).
44. Tsubouchi, A. *et al.* Topological and modality-specific representation of somatosensory information in the fly brain. *Science* **358**, 615–623 (2017).
45. Zinyk, D.L., Mercer, E.H., Harris, E., Anderson, D.J. & Joyner, A.L. Fate mapping of the mouse midbrain-hindbrain constriction using a site-specific recombination system. *Curr. Biol.* **8**, 665-668 (1998).
46. Baumann, O. *et al.* Consensus paper: The role of the cerebellum in perceptual processes. *Cerebellum* **14**, 197-220 (2015).
47. Sweeney, S.T., Broadie, K., Keane, J., Niemann, H. & O'Kane, C.J. Targeted expression of tetanus toxin light chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioural defects. *Neuron* **14**, 341-351 (1995).
48. White, K., Humphrey, D. M. & Hirth, F. The dopaminergic system in the aging brain of *Drosophila*. *Front. Neurosci.* **4**, 205 (2010).
49. Jenett, A. *et al.* A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2**, 991-1001 (2012).
50. Oberdick, J., Baader, S.L. & Schilling, K. From zebra stripes to postal zones: deciphering patterns of gene expression in the cerebellum. *Trends Neurosci.* **9**, 383-390 (1998).

51. Kvon, E.Z. *et al.* Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* **512**, 91-95 (2014).
52. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636-W641 (2019).
53. Visel, A., Minovitsky, S., Dubchak, I., Pennacchio, L.A. VISTA Enhancer Browser - a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88-92 (2007).
54. Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7 (2005).
55. Waterman, M.S., Eggert, M. A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons**.** *J. Mol. Biol.* **197**, 723-728 (1987).
56. Pennacchio, L.A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499-502 (2006).
57. Royo, J.L. *et al.* Transphyletic conservation of developmental regulatory state in animal evolution. *Proc. Natl. Acad. Sci. USA* **108**, 14186-14191 (2011).
58. Kent, W.J. BLAT - the BLAST-like alignment tool. *Genome Res.* **12**, 656-664 (2002).
59. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87-D92 (2019).
60. Urbach, R. A procephalic territory in *Drosophila* exhibiting similarities and dissimilarities compared to the vertebrate midbrain/hindbrain boundary region. *Neural Dev.* **2**, 23 (2007).
61. Butts, T., Green, M.J., Wingate, R.J. Development of the cerebellum: simple steps to make a 'little brain'. *Development* **141**, 4031-4041 (2014).
62. Imai, K.S., Stolfi, A., Levine, M., Satou, Y. Gene regulatory networks underlying the compartmentalization of the *Ciona* central nervous system. *Development* **136**, 285-293 (2009).
63. Pani, A.M. *et al.* Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* **483**, 289-94 (2012).
64. Cao, C., Lemaire, L.A., Wang, W., Yoon, P.H., Choi, Y.A., Parsons, L.R., Matese, J.C., Wang, W., Levine, M., Chen, K. Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349-354 (2019).

65. Ryan, K., Lu, Z., Meinertzhagen, I.A. The CNS connectome of a tadpole larva of *Ciona intestinalis* (L.) highlights sidedness in the brain of a chordate sibling. *Elife* **5**, e16962 (2016).
66. Barsottini, O.G. *et al*. Deafness and vestibulopathy in cerebellar diseases: a practical approach. *Cerebellum* **18**, 1011-1016 (2019).
67. Wray, G.A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
68. Carroll, S.B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
69. Shubin, N., Tabin, C. & Carroll S. Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818-23 (2009).
70. Alberch, P. From genes to phenotype: dynamical systems and evolvability. *Genetica* **84**, 5-11 (1991).
71. Roy, B. et al. Metamorphosis of an identified serotonergic neuron in the *Drosophila* olfactory system. *Neural Dev.* **2**, 20 (2007).
72. Sprecher, S.G. & Hirth, F. Expression and function of the columnar patterning gene *msh* in late embryonic brain development of *Drosophila*. *Dev. Dyn.* **235**, 2920-2929 (2006).
73. Sprecher, S.G., Urbach, R., Technau, G.M., Rijli, F.M., Reichert, H. & Hirth, F. The columnar gene vnd is required for tritocerebral neuromere formation during embryonic brain development of *Drosophila*. *Development* **133**, 4331-4339 (2006).
74. Boll, W. & Noll, M. The *Drosophila Poxn neuro* gene: control of male courtship behaviour and fertility by a complete dissection of all enhancers. *Development* **129**, 5667-5681 (2002).
75. Tautz, D. & Pfeifle, C. A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* **98**, 81-85 (1989).
76. Hirth, F., Hartmann, B. & Reichert, H. Homeotic gene action in embryonic brain development of Drosophila. *Development* **125**, 1579-1589 (1998).
77. Bello, C., Hirth, F. & Gould, A.P. A pulse of the Drosophila Hox protein Abdominal-A schedules the end of neural proliferation via neuroblast apoptosis. *Neuron* **37**, 209-219 (2003).

78. Hassanzadeh, G.H. *et al*. Isolation and characterization of single-chain Fv genes encoding antibodies specific for *Drosophila* Poxn protein. *FEBS Lett.* **437**, 75-80 (1998).
79. Sprecher, S.G. *et al*. Hox gene cross-regulatory interactions in the embryonic brain of Drosophila. *Mech. Dev.* **121**, 527-536 (2004).
80. Walldorf, U. & Gehring, W.J. Empty spiracles, a gap gene containing a homeobox involved in Drosophila head development. *EMBO J.* **11**, 2247–2259 (1992).
81. Faville, R., Kottler, B., Goodhill, G.J., Shaw, P.J. & van Swinderen, B. How deeply does your mutant sleep? Probing arousal to better understand sleep defects in *Drosophila*. *Sci Rep* **5**, 8454 (2015).
82. Mazaud, D. *et al.* Transcriptional regulation of the Glutamate/GABA/Glutamine cycle in adult glia controls motor activity and seizures in *Drosophila*. *J. Neurosci.* **39**, 5269-5283 (2019).
83. Meier-Kolthoff, J.P. *et al.* Complete genome sequence of DSM 30083T, the type strain (U5/41T) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. *Stand. Genomic Sci.* **10**, 2 (2014).

784 **FIGURE LEGENDS**

785

786 **Fig. 1.** The embryonic deutocerebral-tritocerebral boundary gives rise to the antennal
787 mechanosensory motor centre of the adult brain in *Drosophila*. (A, B, D) Confocal
788 images of stage 14 embryonic (A, dorsal; B, lateral) and adult brain (D, frontal)
789 immunolabeled with anti-Brp/nc82; dashed lines demarcate the deutocerebral-
790 tritocerebral boundary (DTB) region; arrowheads indicate antennal mechanosensory
791 motor centre (AMMC). (C) Schematic summarizing gene expression patterns
792 delineating the DTB in the embryonic brain, including the *dachshund* (*dac*) regulatory
793 element R65A11. Abbreviations: PC, protocerebrum; DC, deutocerebrum; TC,
794 tritocerebrum; SOG, subesophaegal ganglion. (E), schematic of adult brain showing
795 AMMC, Wedge and antennal lobes (AL). Scale bars: 20μm (B), 50μm (D).

796

797 **Fig. 2.** FGF8 signalling is required for the formation and maintenance of the embryonic
798 DTB. (A, panels a-c) Confocal images of stage 13/14 embryonic brains immunolabelled
799 for HRP (magenta) and anti-Engrailed (green/yellow). To assist orientation across
800 examples, corresponding axon bundles of developing deutocerebral and tritocerebral
801 sensory nerves (open and close arrows) and their developing sensory neurons (circled)
802 are shown for the wild type (a), *Df(2R)BSC25* (b) and $htl^{AB42-/-}$ (c). These nerves
803 eventually reside in the antenna, which in the imago is equipped with both
804 deutocerebral olfactory neurons and mechanosensory neurons of both deuto-and
805 tritocerebral origin. Dashed yellow lines in a-c indicate the deutocerebral-tritocerebral
806 boundary (DTB). Insets in a-c refer to open rectangles in each panel indicating
807 corresponding normal (a) and affected morphologies of the DTB region (b, c). These
808 include the prefrontal commissure (arrowheads), the width and integrity of

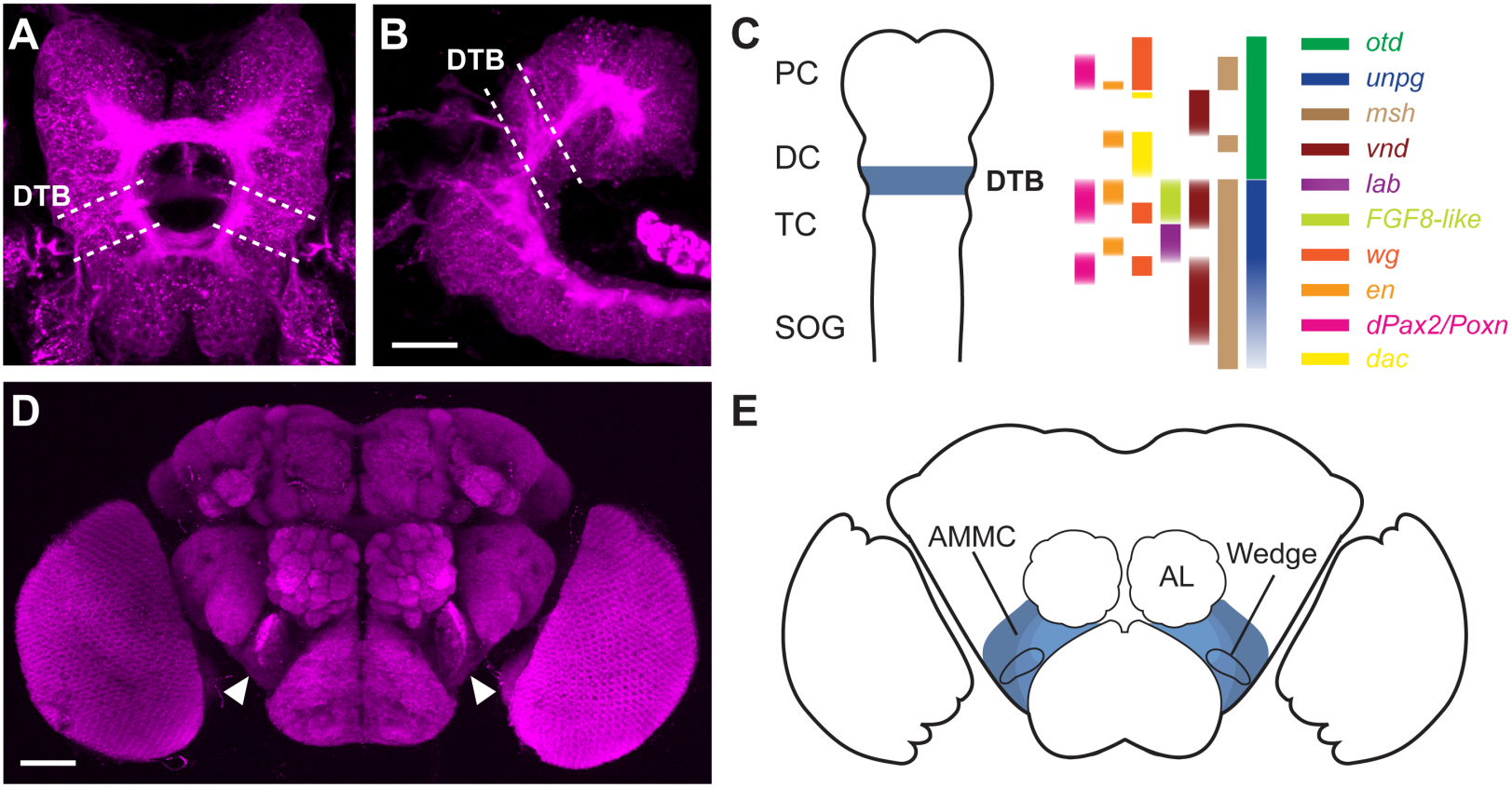
longitudinal connectives (white brackets), and axonal projections at the bifurcation between prefrontal commissure and longitudinal connective (white arrow). All elements are altered in *Df(2R)BSC25* and *htl*$^{AB42-/-}$. Arrowheads and arrow mark the positions where these elements should normally be positioned in the wildtype condition. (Panels d-g) Corresponding color-coded schematics highlight the distribution of engrailed expressing neurons (green); those indicated within yellow rectangle in wildtype (d and removed in the schematic e) are absent in homozygous *Df(2R)BSC25* (b, f) and *htl*$^{AB42-/-}$ (c, g). The morphology of the DTB region in the *Df(2R)BSC25* mutant, a deficiency deleting both *FGF8-like1* and *FGF8-like2* genomic loci, reveals morphological alterations at the DTB (yellow dashed lines) when compared to the wildtype (a, d and e). Likewise, homozygous *htl* null mutant embryos where DTB-specific engrailed expression is lost (c and g) reveal morphological alterations at the DTB (yellow dashed lines) when compared to wildtype (a, d and e). (B) Lateral views of developing embryos at stages E12, E14 and E16. Progressive unpg-lacZ and ems expression patterns reveal successive formation of the DTB. Panels a-c illustrate transgenic unpg-lacZ in control brains at embryonic stages E12, E14 and E16 showing progressive maturation of the DTB domain (bracketed), enlarged in the lower right inset of panel c. Panels d-f. Maintenance of DTB depends on FGF8 signaling: as illustrated by the successive developmental stages of *htl* null mutant embryos (*unpg-lacZ; htl*$^{AB42-/-}$), in which unpg and ems expression patterns are initially visible but those relating to the DTB are subsequently altered by embryonic stage 16. Insets in panels c and f: monochrome enlargements of the open boxed areas in a, f resolve rostro-caudal shortening in the mutant (double arrows insect panel f) compared with the DTB of the wildtype (inset, panel c). HRP immunolabeling is shown as magenta; anti-Ems is shown blue and anti-β-galactosidase shown green. Scale bars: 10μm.

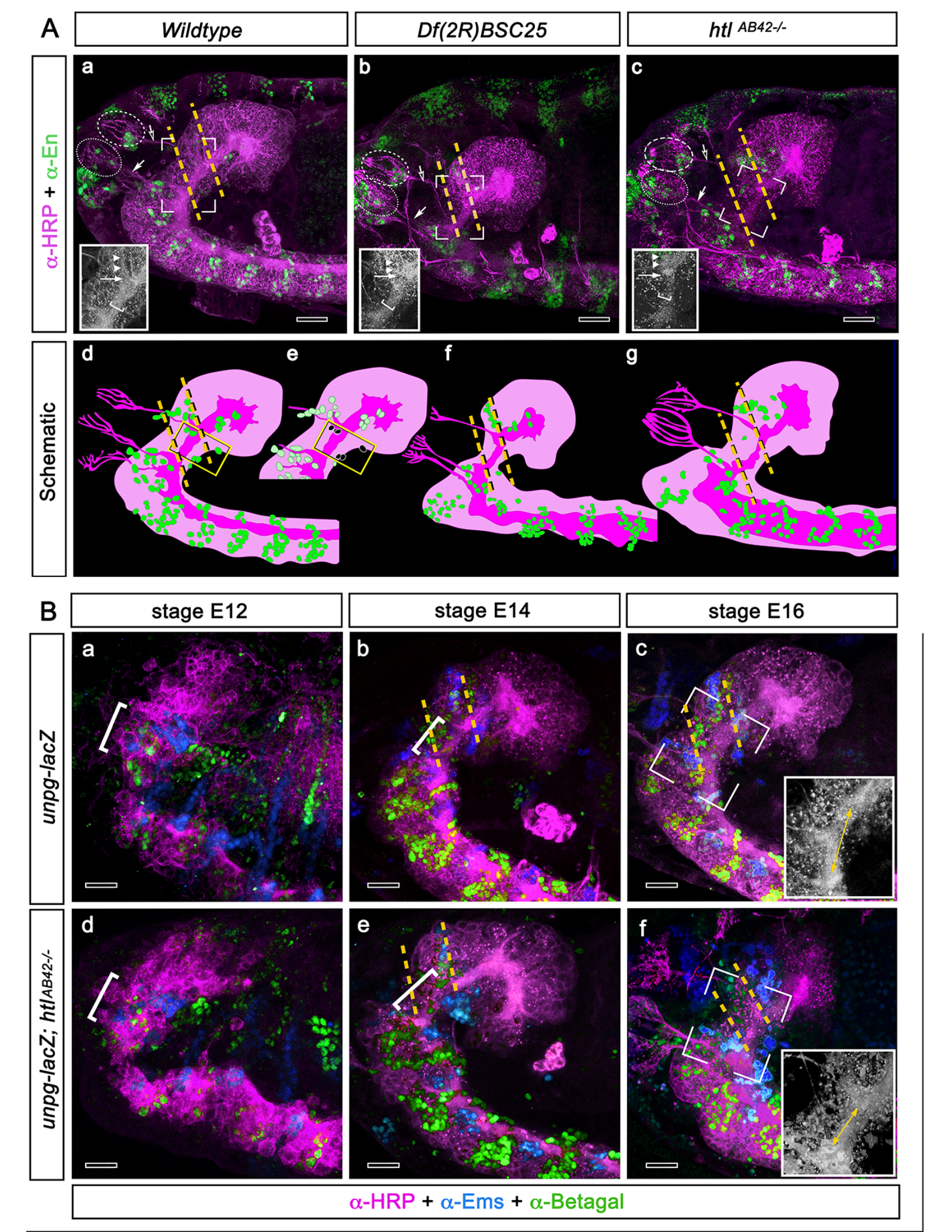
Fig. 3. DTB and AMMC/Wedge-specific expression of the Pax2/5/8 homologue *Pox neuro.* Confocal images of embryonic stage 14 (A-H); anterior is up in A, C, E, G, dorsal views; anterior is to the left in B, D, F, H, lateral views. (A, B)**,** in the anterior embryonic brain, anti-Pox neuro (Poxn) immunolabelling reveals two Poxn expression domains, an anterior at the protocerebral-deutocerebral neuromere boundary and a posterior demarcating the deutocerebral-tritocerebral boundary (DTB) region (arrows, bracket). (C, D), engrailed expression demarcates neuromere boundaries, including the DTB (arrows, bracket). (E, F) *Poxn>mCD8::GFP* expression reveals GFP expression pattern comparable to endogenous Poxn expression (compare with A, B), including the DTB (arrows; bracket). (G, H) *en>nLacZ* expression reveals Engrailed expression pattern comparable to endogenous En expression (compare with C, D), including the DTB that encompasses Poxn expression domain (arrows; bracket). (I-T), confocal images of adult brain; dorsal is up. (I-K) $Poxn^{brain}$*>mCD8::GFP* mediated cell labelling identifies Poxn+ cell clusters (green, arrowheads) in close vicinity to the antennal mechanosensory motor centre (AMMC), the majority of which are anti-Poxn immuno-positive (K, in blue). (L-P) *en>mCD8::GFP* visualises AMMC neurons (arrows) that are located in close vicinity/adjacent to Poxn positive cells (magenta, arrowheads; enlarged views in O, P) that are immunoreactive for anti-GABA (O, blue, arrowheads) like En-expressing cells (P, arrows). (Q-S'), $Poxn^{brain}$*>mCD8::GFP* visualises AMMC/Wedge neurons (arrows) and their projections to antennal glomeruli (Q, Q') to ventrolateral protocerebrum (R-S', middle section of brain), as well as commissural axons of AMMC/Wedge neurons (S', small arrows). (T) *Poxn>tub>mCD8::GFP* mediated genetic tracing of DTB Poxn lineages identifies AMMC/Wedge neurons (arrows). Scale bars: 10μm (K), 20μm (B), 50μm (I).

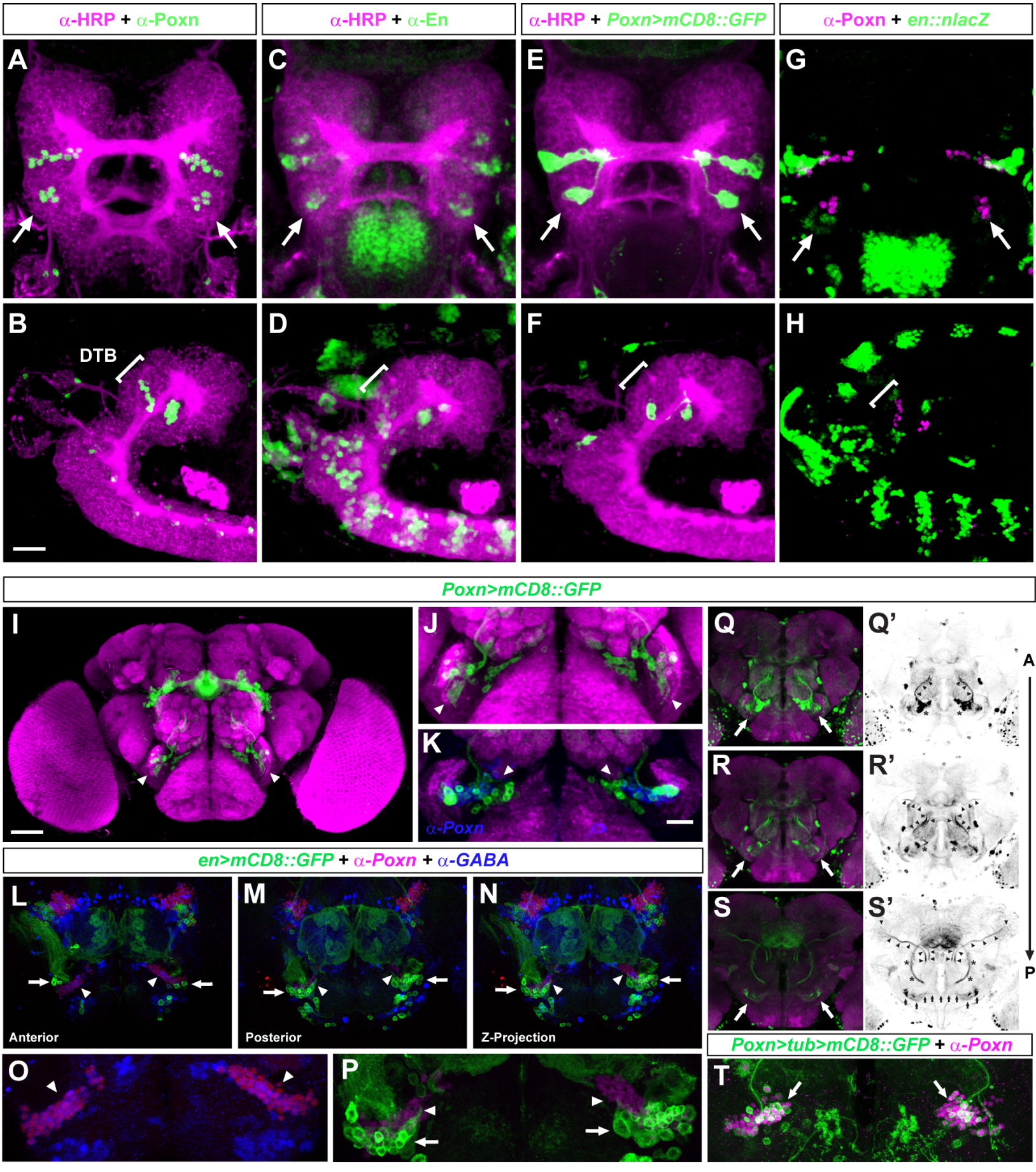
Fig. 4. The AMMC mediates motor coordination in *Drosophila*. (A) Startle-induced negative geotaxis of AMMC-specific GAL4 lines misexpressing *UAS-TNT* and controls (n=150 each). (B) *Drosophila* Arousal Tracking (DART) recording each fly in tube walking back and forth, motors underneath elicit vibration stimuli via digital-to-analogue converter (DAC). (C) *CG9560/dZic-B, R79D08>mCD8::GFP* immunolabeled with anti-Brp/nc82 and anti-GFP visualizes AMMC-specific (bracket) arborizations (arrows); middle panel, rotated brain to depict AMMC-specific projections. (D) Motor behaviour of *R79D08>TNT,* UAS/+ and Gal4/+ control flies; left, raster plots of activity bouts, each lane one individual fly, coloured boxes indicate genotypes; right, stimulus response (main plot) and median response (inset) to repeated mechanical stimulation (dashed orange lines). (E) *inv R88B12>mCD8::GFP* visualizes neuronal projections to AMMC (arrows, enlarged for hemisphere in middle and right panel) encompassed by anti-En (green) and anti-Poxn (magenta) expression domains (bracket). (F) Motor behaviour of *R88B12>TNT,* UAS/+ and Gal4/+ control flies, parameters as in D. (G) *NetA*, *R30A07>mCD8::GFP* expression in AMMC (left, arrows); right, anti-En (red) and anti-Poxn (blue) immunolabelling encompasses *R30A07>mCD8::GFP* domain (brackets) and cells, some co-expressing Engrailed (arrowheads). (H) Motor kinematics of *R30A07>TNT,* UAS/+ and Gal4/+ control flies. (I) *hth R45D07*>m*CD8::GFP* in AMMC (arrows); right, *R54D07>mCD8::GFP* visualises AMMC interneurons and dendritic arborisations close to anti-Engrailed (red) and anti-Poxn (blue) immunolabelled neurons (blue) that encompass AMMC/Wedge (brackets). (J) Motor kinematics of *R45D07>TNT,* UAS/+ and Gal4/+ control flies. Mean ± Standard Error of the Mean (SEM), asterisks, $p<0.05$ (*), $p<0.01$ (**) or $p<0.001$ (***). Scale bars, 50μm (C, E, left) and 10μm (E, right).

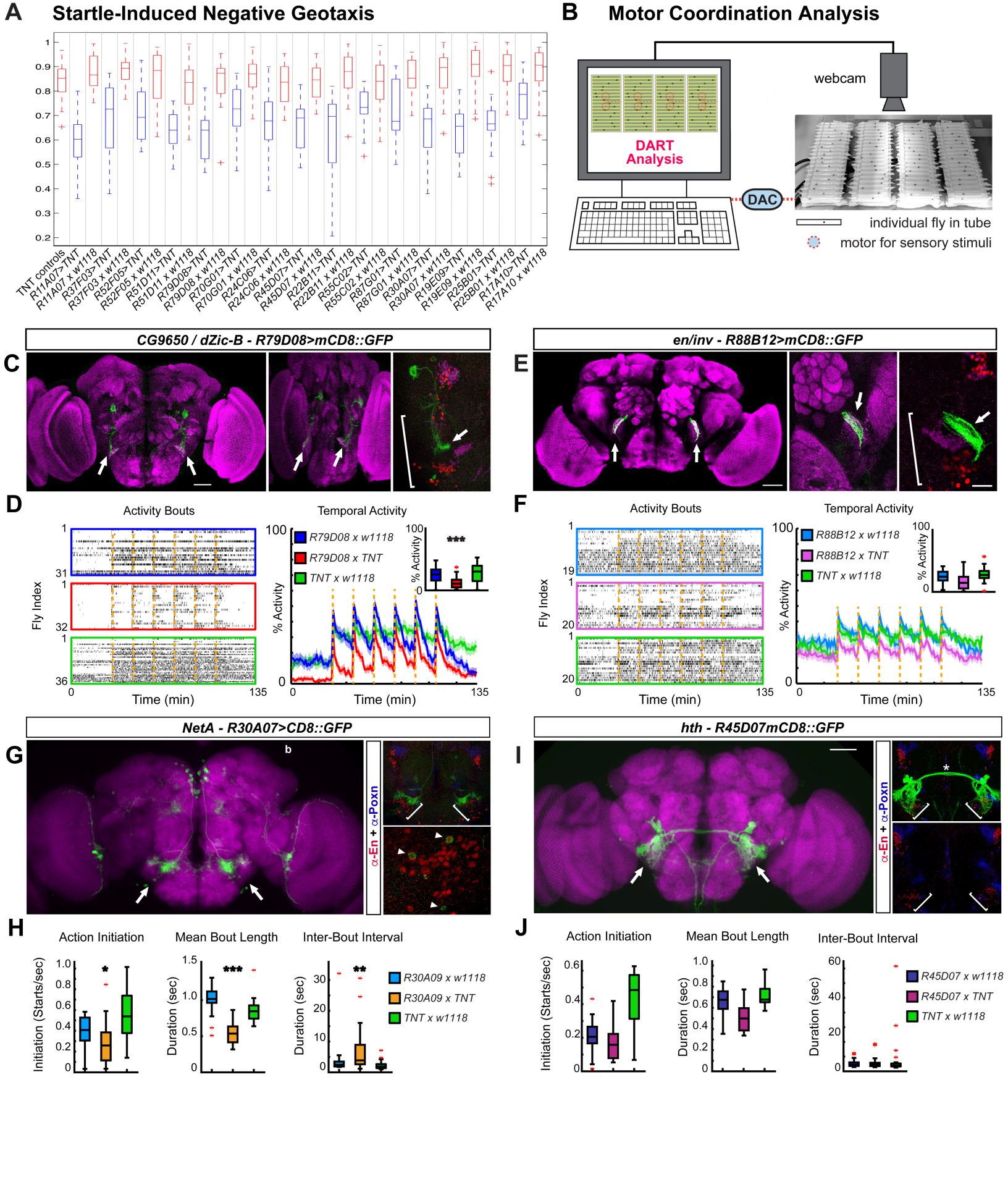
884 **Fig. 5.** Conserved cis-regulatory sequences of *dac/DACH1* direct DTB-AMMC specific
885 expression in *Drosophila* and MHB-specific expression in mouse. (A) Confocal image of
886 stage 10/11 *Drosophila* neuroectoderm of *R65A11-Gal4>mCD8::GFP* embryo (lateral
887 view, anterior left, dorsal up), immunolabeled with anti-GFP (green) and anti-En
888 (magenta); inset illustrates Engrailed expression domains in procephalic
889 neuroectoderm including head spot (hs), antennal spot (as) and intercalary spot (is).
890 *R65A11>GFP* expression (arrow) is seen in DTB primordium (bracket). (B) Confocal
891 image of *R65A11-LexA>mCD8::GFP* expression in AMMC (arrows) of adult *Drosophila*
892 brain, immunolabeled with anti-Brp/NC82 (magenta) and anti-GFP (green). (C) Anti-En
893 and anti-Poxn immunolabeling encompass *R65A11-LexA>mCD8::GFP* in AMMC
894 (bracket). (D) human *DACH1*-specific cis-regulatory sequence (CRE), *hs137,* targets
895 LacZ expression to midbrain hindbrain boundary (MHB, arrow) in E11.5 mouse
896 embryo (VISTA database). (E) Sequence comparison of parts of *D. melanogaster dac*
897 *R65A11*, mouse *mDach1* and human *hDACH1 hs137* CREs. (F) Intragenic locations
898 (black bar, asterisks) of *dac R65A11*, *hDACH1 hs137* and corresponding mouse *mDach1*
899 CRE sequence.
900
901 **Fig. 6.** Phylogenetic comparison of DTB/MHB-related molecular signatures in nervous
902 systems of extant Bilateria. Schematic diagram of homologous gene expression in
903 central nervous system of arthropod *Drosophila melanogaster*, annelid *Platynereis*
904 *dumerilii*, vertebrate *Mus musculus*, ascidian *Ciona intestinalis*, amphioxus
905 *Branchiostoma floridae*, and ectodermal nervous system of hemichordate *Saccoglossus*
906 *kowalevski.* Embryos and bar diagrams are oriented anterior to the left, dorsal up;
907 brown colouring indicates boundary region. Multi-level correspondences of arthropod
908 DTB and vertebrate MHB ground pattern organization suggest ancestral origin of gene

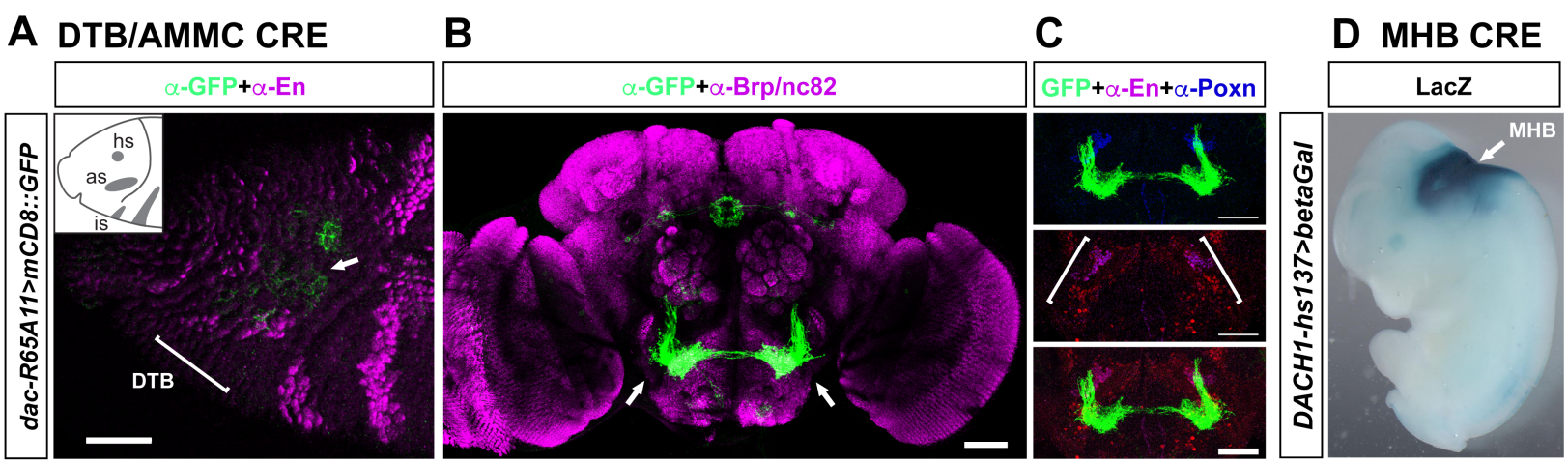
909 regulatory networks (GRNs) for cephalic nervous systems >520 million years ago
910 (Mya) that predate the radiation into protostomes and deuterostomes, and a suggested
911 origin of the MHB-specific and proliferation related isthmic organizer (IsO) function in
912 the vertebrate lineage. Abbreviations: cg, cerebral ganglion; col, collar; cv, cerebral
913 vesicle; dc, deutocerebrum; di, diencephalon; hb, hindbrain; mb, midbrain; n, neck; nc,
914 nerve cord; pc, protocerebrum; pr, proboscis; sg, segmental ganglia; sv, sensory vesicle;
915 tc, tritocerebrum; tel, telencephalon; tr, trunk.


A
DTB
B
DTB
C
PC
DC
TC
SOG
DTB
otd
unpg
msh
vnd
lab
FGF8-like
wg
en
dPax2/Poxn
dac
D
E
AMMC
AL
Wedge


A
Wildtype
Df(2R)BSC25
htl AB42-/-
α-HRP + α-En
a
b
c
Schematic
d
e
f
g
B
stage E12
stage E14
stage E16
unpg-lacZ
unpg-lacZ; htlAB42-/-
a
b
c
d
e
f
α-HRP + α-Ems + α-Betagal


α-HRP + α-Poxn
α-HRP + α-En
α-HRP + Poxn>mCD8::GFP
α-Poxn + en::nlacZ
A
B
DTB
C
D
E
F
G
H
Poxn>mCD8::GFP
I
J
K
α-Poxn
Q
Q'
R
R'
S
S'
A
P
en>mCD8::GFP + α-Poxn + α-GABA
L
Anterior
M
Posterior
N
Z-Projection
O
P
Poxn>tub>mCD8::GFP + α-Poxn
T

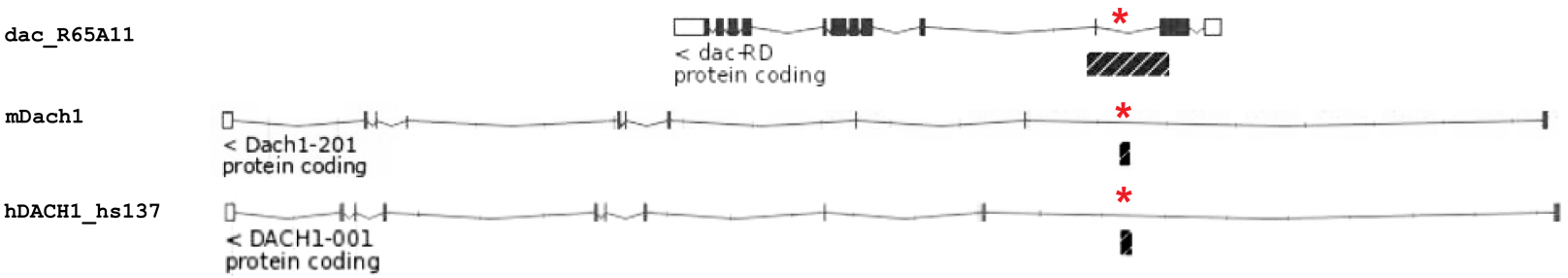

A
Startle-Induced Negative Geotaxis
TNT controls
R11A07>TNT
R11A07 x w1118
R37F03>TNT
R37F03 x w1118
R52F05>TNT
R52F05 x w1118
R51D11>TNT
R51D11 x w1118
R79D08>TNT
R79D08 x w1118
R70G01>TNT
R70G01 x w1118
R24C06>TNT
R24C06 x w1118
R45D07>TNT
R45D07 x w1118
R22B11>TNT
R22B11 x w1118
R55C02>TNT
R55C02 x w1118
R87G01>TNT
R87G01 x w1118
R30A07>TNT
R30A07 x w1118
R19E09>TNT
R19E09 x w1118
R25B01>TNT
R25B01 x w1118
R17A10>TNT
R17A10 x w1118
B
Motor Coordination Analysis
webcam
DART Analysis
DAC
individual fly in tube
motor for sensory stimuli
CG9650 / dZic-B - R79D08>mCD8::GFP
C
D
Activity Bouts
Temporal Activity
Fly Index
Time (min)
% Activity
R79D08 x w1118
R79D08 x TNT
TNT x w1118
***
en/inv - R88B12>mCD8::GFP
E
F
R88B12 x w1118
R88B12 x TNT
TNT x w1118
NetA - R30A07>CD8::GFP
G
b
α-En + α-Poxn
hth - R45D07mCD8::GFP
I
H
Action Initiation
Mean Bout Length
Inter-Bout Interval
Initiation (Starts/sec)
Duration (sec)
R30A09 x w1118
R30A09 x TNT
TNT x w1118
J
R45D07 x w1118
R45D07 x TNT
TNT x w1118

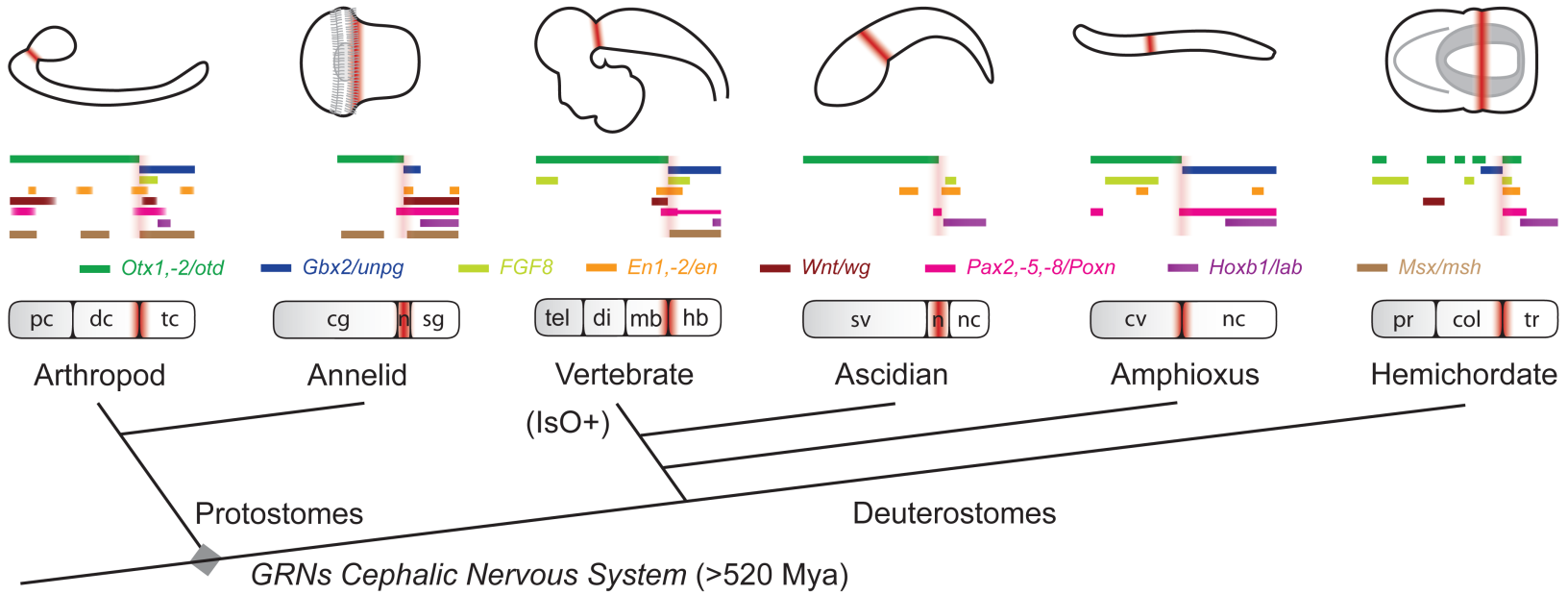


E CRE sequence comparison

```
dac_R65A11      -------CGAAAATATAGCCCTTTTTCACGCTCTATTTATAGCATTCACATTCTTTCTTTTTTTTTGCACTTTT---TAGCTGGCATATCCTTTCGACTTCCGCCATTCGAGGCTCGCCC
DACH1_hs137     GCAATTTTGAAAAAGAAAACAATGGTTAGGAAGACTTTATTTTATTCTAATTATATTTGATACTTTAAATGTTTATCTATCTAATATATCTCTTATTAAACAATTCTTTAGCTCTGGTAT
mDach1          ---------------------------------ACCCTATTTTATTCTAATTGTATTCGGTAGTTAAAATGTTTACCTATCTCCATAATATCTTATTAAACAGCACTCCAGTGATAGACA
                                                  ***   ****  *** * *    *  **   *  ***   ** **     **   **      *     *       * *

R65A11_dac      AATTTCCGTTTCGAGTTTAATTAATTTAATAAACAAATTCTTTTCGCTCTAAAAACTCTCAAGTG--TATCGATACGATGCGTTTCTTTTTTTCCTTCGTTAAATAAATAATAACCAAAA
DACH1_hs137     ATTTTTTAGTAAAGTTTTAATAACTTTC---------------TCAGAATTAAAAATAGCATATGATATCAGAG----ACCAA-------------------------CAATAAAAGTAA
mDach1          --TTTTTGATAAAGTTATAATAACTTTT---------------TTATATTAAGAATAGAAAAATAACAACAGAGACCAATAGT-------------------------TATTAAAAGTAA
                  ***    *     * **** * ***                *     * * **     *  *       **                                      * ***    **

dac_R65A11      AAAAAAAAAACCAAAAAGTAGGAGGAGAAAAGTTATTGCCATAGTTTTTTTATTATACTTGTGTGTTTACCTTTCTGGTGGCTTGATCGATAGGCATCTGCAATTAAAAAGAGAAGAAGA
DACH1_hs137     TCGA-ATTTAGAGAAATGTTTAGAAAATAAATAGATCTATACATATTTTGTTAAAGACCAATCAATTGAT-CTGCTGTTTTCTATGTA------------AAATGAAACAAT--GCTTAT
mDach1          TAAA-ATTTACTGAAATATATAGAAAATAAAGAGATTCACATACATTTGGCTTAGGACTTATCAGTTTAT-TTGCTGGTGTTTAAGTA------------AAGCGAAGCAAT--GGCTTC
                   * *   *   ***  *      *  ***   **    * *  ***        **   *   ** *   * *** *   *   *             *   **  *

dac_R65A11      AGAGACAAGTGAGGCAAAATTGTTAA----------ACGTTTTGTGTAAGCTTTAATACGAAAAACAAGTACTGCAACATAACGGAAGGA-AACACGGCTTAAATTCGGGGCACAAATGC
DACH1_hs137     TATTAAAATATAGGCTGAATTCTTTGCCTGATTTTCCGCTTTTGTAAAATGCATTGCTTGAGAAACAAGTCTTCAAAAAACAAGGCCTA--AAAAAATCTACTACTCAGGGTGCCTTTGA
mDach1          CGATTAAATGTAGGCGAAGTTCTTTGCCTGGT-TTTTGCTTTTGTAAAATGCATTGCAAGAGAAACAAGTCTTCAAAAAACAGGGCCTAAAAAAAAATCTACTACTCAGGGTGCCTTTGA
                      **   ****  * ** **              ******  **    *     ** ********  *  ** *  * **      ** *   **   * ** ***  *   **

dac_R65A11      TGAAAGGGAAGTTT--TTCATTGACGGGTTCGTTCTGACGGACTTGCATTTTGGCGGG----------CAAGCGGGTGTGAAAATGCACACG----------CCCCGAGAACCCCCCTTT
DACH1_hs137     GGATATTTAATTAAAATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATGGATTCTCATTTGCATATTGCATACAATTCATCAATTACTTCAAGTAT
mDach1          GGATATTTAATTAAAATCTAATCCTGCATTCATTAAGGCTCACATAAATTAAGCTGTCATTCATAAGATTTATGGATTCTCATTTGCATATTGCATACAATTCATCAATTACTTCAAGTAT
                 ** *   ** *    *  * *   *  *** **  * *  ** *  ***  *  *                  ** *    *  **** *           *  * *  ** *   * *

dac_R65A11      CCACCCCCCCTGGACCCCTTTATCCAGCCCACTGGCCAAAAACAATTTGTAATTATCCACAGAGAGCGCTGCCTTCAGCGGTTTCGCATTTCCCCTTTCGCTCGCTCTCCCAACTTGTTT
DACH1_hs137     GAAAGGAGCACATTTCCCTTGGAGCTGCCTGCTACCCTGCCAACATTTGAAATGAGGGAAAGAG--CAAGACTGTCAGGCATTCACACAAACTTTCTTCCAAATGTCTG-CTCCTTGATT
mDach1          GAAAGGAGCACATTTCCCTTGGAGCTGCCTGCTACCCTGCCAACATCTGAAATGAGGGAAAGAG--CAAGACTGTCAGGCATTCACACAAACTTTCTTCCAAATGTCTG-CTCCTTGATT
                  *     *      *****    * ***  **  **    *  ** ** *** *   * ****  *    *  ****   **        *    ***      ***  *  **** **

dac_R65A11      CAATTTAGCGCAAAACTTTTTCAACCTAATAATAGGTTTAACCGCATTTTTAACCGTTCCTCATGTTCGGTCCGGTTCGGTTTTCAAAACCGGGAATCGTACTTAGACTGGGTCTCCTTA
DACH1_hs137     AATCTAATTTTCTAGATCTTCCCTAC---AAGATACACCAACAGCCCTGGTGCACATTTCTATTATTTCGCCTGTCTTCATCACCAAAATCT---TCCAGTCTTTGAATACTTCTTCATG
mDach1          AATCTAATTTTCTAGATATTCCCTAC---AAGATATAACAACAGCTCT-AAGCACATTTCTATTATTTCTCCTGTCTTCATCACCAAAACCC---TCTAGTCTTTGAATACCTCTTCATG
                 *  * *      *  * ** *   *    *        *** **  *      * ** **  * **    * *  *   *   ***** *         *** ** *   *** * *

dac_R65A11      TTTCTGTTCTGGCTCTCTGTACAACTTTTCATTGAGAAAAATGTAACTAGTTTTTCATAGCAACGGAATACAATTTAATCCAATAATCCAATAGTTTAATCCAATACAAATGATATTACT
DACH1_hs137     TCTGAGTGATTATTCT-----------TTATTTGCTAATAA--------------------TATGGCATTGATGATATTTCCCTATTCAAATGTGTTTGATATGTACAAAGATGTAGGTT
mDach1          TCTGAGTGATTATTAT-----------TTATTTGCTAATGA--------------------TATGGCATTGATGCTTTCTTGATGATAAAATGGCTTTAATGTGTACAAATAAGAAAGCT
                * *  **  *   * *           **  ***  **  *                     * ** **  *   *       *  *  ***   **       ******        *
```

F Topology and intragenic location of CRE




Otx1,-2/otd
Gbx2/unpg
FGF8
En1,-2/en
Wnt/wg
Pax2,-5,-8/Poxn
Hoxb1/lab
Msx/msh
pc
dc
tc
cg
n
sg
tel
di
mb
hb
sv
n
nc
cv
nc
pr
col
tr
Arthropod
Annelid
Vertebrate
Ascidian
Amphioxus
Hemichordate
(IsO+)
Protostomes
Deuterostomes
GRNs Cephalic Nervous System (>520 Mya)